Systematic Entomology (2007), **32**, 420–428 DOI: 10.1111/j.1365-3113.2007.00385.x

Double trouble for grasshopper molecular systematics: intra-individual heterogeneity of both mitochondrial 12S-valine-16S and nuclear internal transcribed spacer ribosomal DNA sequences in *Hesperotettix viridis* (Orthoptera: Acrididae)

GREGORY A. SWORD[1,2], LAURA B. SENIOR[1], JOHN F. GASKIN[1] and ANTHONY JOERN[3]

[1]Northern Plains Agricultural Research Laboratory, United States Department of Agriculture, Agricultural Research Service, Sidney, Montana, U.S.A., [2]School of Biological Sciences, The University of Sydney, NSW, Australia. [3]Division of Biology, Kansas State University, Manhattan, Kansas, U.S.A.

Abstract. *Hesperotettix viridis* grasshoppers (Orthoptera: Acrididae: Melanoplinae) exhibit intra-individual variation in both mitochondrial 12S-valine-16S and nuclear internal transcribed spacer (ITS) ribosomal DNA sequences. These findings violate core assumptions underlying DNA sequence data obtained via polymerase chain reaction (PCR) amplification for use in molecular systematics investigations. Undetected intra-individual variation of this sort can confound phylogenetic analyses at a range of taxonomic levels. The use of a DNA extraction protocol designed to enrich mitochondrial DNA as well as an initial long PCR of approximately 40% of the grasshopper mitochondrial genome failed to control for the presence of paralogous mitochondrial DNA-like sequences within individuals. These findings constitute the first demonstration of intra-individual heterogeneity in mitochondrial DNA-like sequences in the grasshopper subfamily, Melanoplinae, and only the second report of intra-individual variation in nuclear ITS ribosomal DNA sequences in grasshoppers. The fact that intra-individual variation was detected in two independent DNA marker sets in the same organism strengthens the notion that the orthology of PCR-derived DNA sequences should be examined thoroughly prior to their use in molecular phylogenetic analyses or as DNA barcodes.

Introduction

The rapid evolution and ease of polymerase chain reaction (PCR) amplification of both mitochondrial genes (mtDNA) and internal transcribed spacer regions (ITS1 and 2) of nuclear ribosomal DNA (rDNA) have facilitated the use of these sequences as genetic markers in numerous population-level phylogeographical and species boundary investigations (Hillis & Dixon, 1991; Simon *et al*., 1994; Hillis *et al*., 1996; Avise, 2004). Two important assumptions underlying the use of PCR amplification and direct DNA sequencing in phylogenetic analyses are that: (1) the PCR product obtained is the desired target sequence, and (2) all of the amplified DNA sequences represent a single orthologous locus that is invariant within individuals, but homologous due to common ancestry with sequences obtained from other individuals (Palumbi, 1996; Sanderson & Shaffer, 2002). A growing number of examples indicate that these assumptions are routinely violated due to the presence of intra-individual variation in DNA sequences that share homology as a result of gene duplication events rather than common ancestry. If undetected, the amplification and sequencing of these paralogous loci can potentially confound molecular phylogenetic analyses (Sorenson & Quinn, 1998; Bensasson *et al*., 2001a; Álvarez & Wendel, 2003; Funk & Omland, 2003).

Correspondence: G. A. Sword, School of Biological Sciences, The University of Sydney, Macleay Building A12, Sydney, NSW 2006, Australia. E-mail: greg.sword@bio.usyd.edu.au

Intra-individual variation in mtDNA sequences can arise due to mitochondrial heteroplasmy or gene duplications (Bensasson *et al*., 2000, 2001a), although there is little evidence to suggest that such events have influenced the outcome of phylogenetic analyses (Rokas *et al*., 2003; Ballard & Whitlock, 2004; Thalmann *et al*., 2004). However, another source of intra-individual variation that has received considerable attention of late is the integration of mtDNA into the nuclear genome. This cytoplasmic transfer of organelle DNA results in nuclear-mitochondrial pseudogenes (numts) (Lopez *et al*., 1994) that have been found in a variety of animal species (Zhang & Hewitt, 1996; Bensasson *et al*., 2001a; Richly & Leister, 2004). Numts appear to be particularly common in the Orthoptera, especially among grasshoppers (family: Acrididae). Grasshopper numts can exist in high copy numbers and have been identified in at least 14 different grasshopper species representing five different subfamilies, the Calliptaminae, Cyrtacanthacridinae, Gomphocerinae, Oedipodinae and Podisminae (Zhang & Hewitt, 1996; Bensasson *et al*., 2000, 2001a).

When encountered, one way to deal with a mtDNA paralogy problem is to select an independent marker. The ITS region of nuclear ribosomal genes is also popular for use in phylogenetic analyses at lower taxonomic levels in many taxa, including insects (Loxdale & Lushai, 1998; Wörheide *et al*., 2004). In the eukaryotic genome, the nuclear ITS regions are part of a larger rDNA array that exists typically in several hundred tandemly repeated copies (Hillis & Dixon, 1991). The nucleotide sequences of these arrays are thought to be homogenized within populations by concerted evolution (Liao, 1999; Avise, 2004). Despite this, intragenomic variation in ITS sequences has been found in several different species, including grasshoppers (Vogler & Desalle, 1994; Leo & Barker, 2002; Parkin & Butlin, 2004; Wörheide *et al*., 2004).

The widespread occurrence of intragenomic variation has prompted some investigators to suggest that the orthology of PCR products should always be tested and not assumed (Dowling *et al*., 1996). Here we demonstrate that the orthology assumption is violated in *Hesperotettix viridis* (Orthoptera, Acrididae, Melanoplinae) grasshoppers for both 12S-valine-16S mtDNA and nuclear ITS rDNA sequences. This example of intra-individual heterogeneity in two independent DNA marker sets within the same species provides further support for the notion that the orthology of PCR-amplified DNA sequences should not be an a priori assumption in molecular applications such as phylogenetics or DNA barcoding.

Materials and methods

This report stems from an attempt to develop 12S-valine-16S mtDNA and nuclear rDNA ITS sequences as markers for use in an analysis of host plant-associated genetic differentiation in *H. viridis*. Initial attempts to sequence both gene regions directly from PCR products obtained using whole genomic DNA extractions in conjunction with published universal primers resulted in evidence of multiple DNAs amplified during PCR. Peaks under peaks, suggesting the presence of more than one type of base at a given site, were evident in the sequence chromatograms, some substantial enough to result in unresolved base calls and occasionally unresolved reads over larger regions. As a result, we set out to test the orthology hypothesis and screened for the presence of heterogeneous DNAs in our PCR products using a clone and sequence approach. In the case of mtDNA, we also used the same cloning approach to examine the efficacy of a nested long PCR technique in which a large ~6000 bp portion of the ~15 700 bp grasshopper mitochondrial genome (Flook *et al*., 1995) was amplified initially in an attempt to control for the presence of numts (Roehrdanz & Degrugillier, 1998; Bensasson *et al*., 2001a).

DNA extractions

DNA extractions were conducted on five *H. viridis* individuals hereto referred to as A, B, C, D and E. Individuals A, B, C, D and E correspond to insects AS-3, KS-2, KG-1, KS-3 and AS-4, respectively, in the amplified fragment length polymorphism (AFLP) analysis of Sword *et al*. (2005), but were used first in this study without prior knowledge of the relationships among populations. A hind femur of each grasshopper was ground with 20 mg of sterile white quartz sand in a 1.5 mL microfuge tube using disposable plastic pestles. DNA extractions were conducted using Qiagen DNeasy tissue kits following the animal tissue protocol (Qiagen, Valencia, California, U.S.A.). An additional extraction designed to enrich mtDNA was conducted with specimen C using the alkaline lysis protocol of Tamura & Aotsuka (1988) with 32 mg of thorax and leg muscle tissue. Extractions were stored at –20 °C.

mtDNA PCR

The 12S-valine-16S mtDNA region was amplified from insects A, B and C using universal mtDNA primers modified for consistency with the *Locusta migratoria* mitochondrial genome (16Sa, LR-J-13417: 5′-ATGTTTTT-GATAAACAGGCG and 12Sc, Sr-N-14275: 5′-AAGGTGGATTTGATAGTAAT) (Simon *et al*., 1994; Flook *et al*., 1995). Template DNA used for individual A was that extracted according to Tamura & Aotsuka (1988) to examine the effectiveness of this protocol in enriching mtDNA and potentially eliminating numts. PCR reactions contained 1 μL genomic DNA at extracted concentration, 0.2 mM dNTPs, 0.75 μL (2.5 U) AmpliTaq DNA polymerase (Applied Biosystems, Foster City, California, U.S.A.), 1 × GeneAmp PCR buffer (Applied Biosystems), 1 μM each primer and milliQ water in a final volume of 100 μL. Cycling conditions were 94 °C for 5 min; 35 cycles of 1 min at 94 °C, 1 min at 48 °C and 1 min at 72 °C; and a final extension time of 5 min at 72 °C. PCR products were screened for expected size (~940 bp) under ultraviolet light on agarose gels stained with ethidium bromide, then cloned as described below.

mtDNA nested long PCR

We employed a long PCR to amplify ~40% (~6000 bp) of the mitochondrial genome of the same individuals screened above (A, B and C) to determine if it could isolate mtDNA, thereby eliminating suspected numt contamination. The primers were from Roehrdanz & Degrugillier (1998), but modified slightly to conform to the mtDNA sequence of *Locusta migratoria* (Flook *et al*., 1995). They were 12S (5′-AGACTAGGATTAGATACCCTATTAT) and N4 (5′-GGAGCTTCAACATGAGCCTT) using the notation of Roehrdanz & Degrugillier (1998). The Expand Long Template PCR system was employed following the manufacturer's system 3 protocol (Roche Diagnostics, Basel, Switzerland). The reactions contained 125 ng of template DNA derived from the *H. viridis* whole genomic DNA extractions, Expand Long Template buffer 3 (10× concentration with 27.5 mM $MgCl_2$), an additional 1.25 mM $MgCl_2$, 0.5 mM dNTPs, 300 nM each primer, 0.75 μL (3.75 U) Expand Long Template Enzyme mix, and milliQ water in a final volume of 50 μL. Cycling conditions were 92 °C for 2 min; ten cycles of 10 s at 92 °C, 30 s at 55 °C and 13 min at 68 °C; 20 cycles of 10 s at 92 °C, 30 s at 55 °C and 13 min at 68 °C +20 s for each successive cycle; and a final extension time of 7 min at 68 °C. The ~6000 bp long PCR products were visualized under ultraviolet light on agarose gels stained with ethidium bromide and gel extracted using the GENECLEAN Turbo Nucleic Acid Purification Kit (Qbiogene, Carlsbad, California, U.S.A.).

For each sample, the purified long PCR putative mtDNA products from each individual were used as a template for a subsequent PCR using the universal mtDNA primers and reaction conditions as described above. The resultant ~940 bp products were gel purified using the QIAquick Gel Extraction Kit (Qiagen) and then cloned as described below.

ITS PCR

The ITS1-5.8S-ITS2 nuclear rDNA regions of specimens D and E were amplified using eukaryote-specific primers as per Weekers *et al*. (2001) (5′-TAGAGGAAG-TAAAAGTCG and 5′-GCTTAAATTCAGCGG). PCR reactions contained 1 μL of genomic DNA, and contained 0.2 mM dNTPs, 1.25 U AmpliTaq Gold polymerase (Applied Biosystems), 1× GeneAmp PCR Gold buffer (Applied Biosystems), 1% dimethylsulphoxide (DMSO), 0.5 μM each primer and milliQ water to a final volume of 50 μL. Cycling conditions were 95 °C for 10 min; 35 cycles of 30 s at 95 °C, 30 s at 55 °C and 1 min at 72 °C; and a final extension time of 7 min at 72 °C. The PCR products were gel screened for appropriate size as described above and cloned as below.

Cloning and sequencing

The cloning of both mtDNA and nuclear rDNA was in accordance with the manufacturer's instructions for the TOPO TA Cloning Kit for Sequencing (Invitrogen, Carlsband, California, U.S.A.). PCR products were ligated into pCR®-TOPO plasmids (Invitrogen). These plasmids were used subsequently to transform One Shot® TOP10 Electrocomp *Escherichia coli* (Invitrogen). Clones were incubated overnight at 37 °C on LB plates with 50 μg/mL kanamycin. For each specimen, multiple colonies were picked and separately cultured overnight in LB broth containing 50 μg/mL kanamycin. The plasmids were prepared using the QIAprep Spin Miniprep Kit (Qiagen) in accordance with the manufacturer's instructions. Prior to sequencing, all clones were screened via restriction digest analysis to confirm the presence of an insert of appropriate size.

Sequencing was accomplished using a Beckman CEQ2000XL automated DNA analysis system (Beckman Coulter, Fullerton, California, U.S.A.). The manufacturer's protocol was followed for sample preparation and dye terminator cycle sequencing. For individuals A, B and C (those amplified only with universal mtDNA primers), four, eight and eight clones, respectively, were sequenced (GenBank/EMBL ascension numbers EF210398–EF210401, EF210376–EF210383 and EF210416–EF210423). These same individuals were used in the nested PCR (long PCR followed by PCR using universal primers) and 14 clones of each were sequenced (GenBank/EMBL ascension numbers EF210402–EF210415, EF210384–EF210397 and EF210424–EF210437). More clones were sequenced following nested long PCR than standard PCR based on the expectation that divergent sequences would be rare, if not absent, and require a greater sampling effort to detect. For individuals D and E (those amplified with ITS primers), ten and nine clones, respectively, were sequenced (GenBank/EMBL accession numbers EF213122–EF213140). All clones were sequenced in both directions.

Analysis

Sequences were edited, aligned, and visually optimized in SEQUENCHER 4.1.2 (Gene Codes, Ann Arbor, Michigan, U.S.A.). Within individual pairwise sequence divergences (uncorrected P distances) among clones were calculated in PAUP 4.10b10 (Swofford, 2002). Intragenomic sequence variation above that attributed to *Taq* polymerase error was tested for by assessing the goodness of fit between the expected and observed number of changes using the binomial test (Sokal & Rohlf, 1995). The expected number of *Taq* polymerase errors was calculated based on a *Taq* error rate of 7.3×10^{-5} errors/bp/duplication (Kobayashi *et al*., 1999). In our 12S-valine-16S amplifications, the expected number of bp changes per individual assuming a maximum 937 bp target sequence in a 35 cycle PCR was 2.4. For the ITS amplifications, the expected number of changes per individual was 2.03 assuming a target sequence of a maximum 795 bp in a 35 cycle PCR. Because an additional PCR step was introduced in our mtDNA nested long PCR procedure, the probability of errors introduced by *Taq* error should be higher. However, due to the high fidelity *Taq* used

in long PCR with an error rate of 4.8×10^{-6} errors/bp/duplication (Roche Diagnostics), the expected number of errors introduced during this initial PCR was only 0.13 changes per reaction. Adding this to the number of expected changes in the second PCR results in a total of 2.53 (rounded to 3) expected changes following both bouts of PCR. Based on these error rates, the probability of errors per bp was calculated for each PCR using the maximum observed sequence length and then used as the expected frequency of changes in each binomial test.

BLAST searches (http://www.ncbi.nlm.nih.gov/BLAST/) were conducted on highly divergent sequences to confirm grasshopper origin. Haplotype networks for sequenced clones from each gene region were estimated using statistical parsimony (95%) in the software TCS (Clement *et al*., 2000) using the genealogical reconstruction algorithms of Templeton *et al*. (1992). Some clonal haplotypes were too distantly related to be connected to the network at the 95% parsimony level, and in these cases the haplotypes were connected by dashed lines to the perimeter drawn around the network.

To identify potentially nonfunctional sequences among our mtDNA clones, all cloned sequences were screened for polymorphism at a number of phylogenetically conserved motifs. Conservative motifs of ≥ 5 bp were identified in an alignment of 12S-valine-16S sequences from related grasshopper species representing five genera from two different subfamilies outside of the Melanoplinae obtained from a BLAST search using the most commonly obtained haplotype from individual B (sequence B-2a in Fig. 1A). Taxa used in the alignment and their respective GenBank accession numbers were as follows: Cyrtacanthacridinae: *Schistocerca gregaria* (AY605952), *Cyrtacanthacris tatarica* (AY605953), *Valanga* sp. (AY605956), *Acanthacris ruficornis*

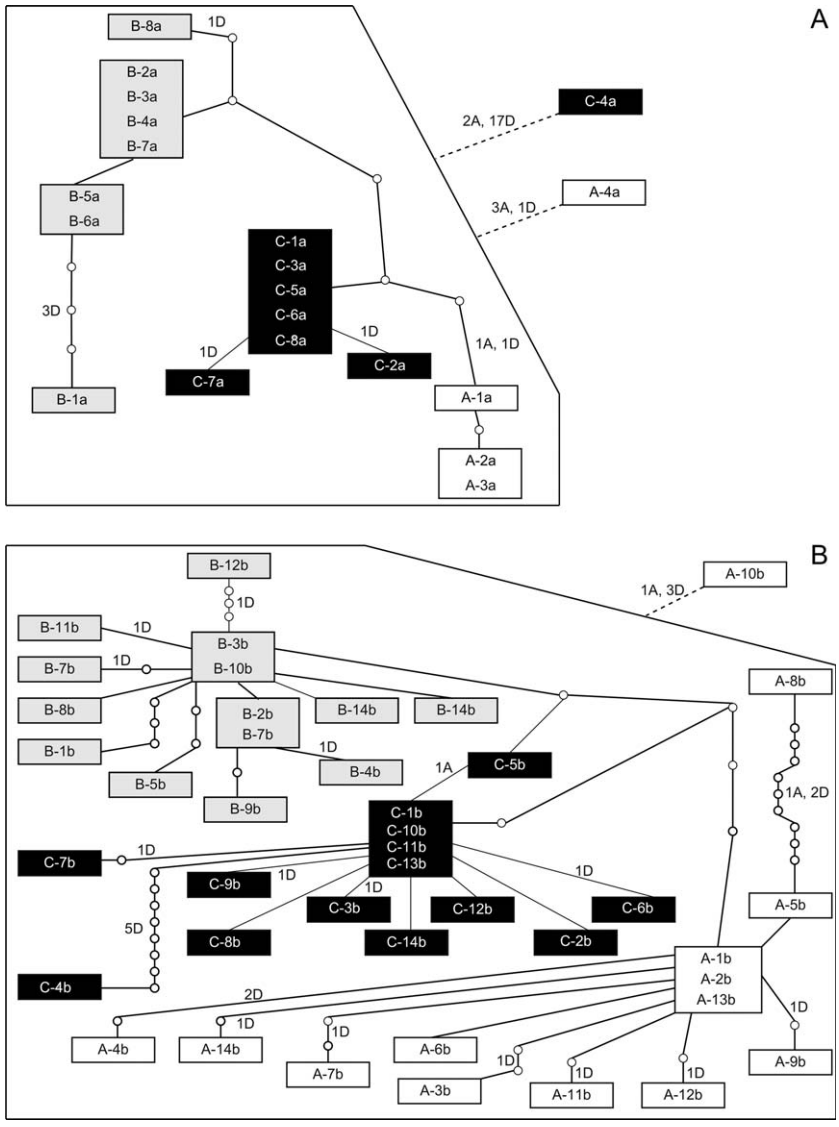

Fig. 1. A, Haplotype network for the 12S-valine-16S mitochondrial DNA (mtDNA) region for 20 sequenced clones from three individual *Hesperotettix viridis* grasshoppers (A, B and C). Squares represent haplotypes recovered, and circles along lineages in between squares indicate haplotypes not recovered. Each link between haplotypes indicates one mutational event, with indels, no matter what size, coded as a single event. The angle of bifurcation and the length of the link between haplotypes have no significance. The perimeter that surrounds the haplotype network indicates haplotypes that could be connected within the limits of parsimony (95%). Haplotypes that could not be connected to the network are connected to the perimeter by a dashed line. Cloned haplotypes found to contain mutations in highly conserved motifs are annotated with the number of changes and their respective character states (A = ancestral; D = derived). B, Haplotype network as described above for the 12S-valine-16S mtDNA region for 42 sequenced clones from the same three individuals following an initial long polymerase chain reaction amplification of ~40% of the mitochondrial genome in an attempt to control for the presence of nuclear-mitochondrial pseudogenes.

(AY605954), *Nomadacris succincta* (AY605955); Oedipodinae: *Locusta migratoria* (X80245). Polymorphic bases or indels among clones within these conserved regions were identified and their ancestral vs. derived state relative to the conserved motif was noted.

Results

Intra-individual heterogeneity in mtDNA sequences

Standard PCR amplification of the 12S-valine-16S mtDNA region in individuals A, B and C yielded PCR products that varied considerably within individuals (Table 1; Fig. 1A). Overall, pairwise divergences of sequences within individuals ranged from 0 to 5.0% and varied in length from 931 to 937 bp. Intra-individual heterogeneity among the recovered clones could not be explained by *Taq* polymerase errors during PCR for any of the individuals (Table 1). Intra-individual variation was observed regardless of whether the initial template DNA used in the PCR amplifications came from whole genomic DNA extractions (individuals B and C) or an alkaline lysis extraction protocol designed for the isolation of mtDNA (individual A). Gene genealogy analysis of the 20 sequences obtained from the three different individuals revealed that some of the sequences were not monophyletic with respect to the individual genomes from which they were recovered (Fig. 1A).

Long PCR amplification of approximately 40% of the grasshopper mitochondrial genome from whole genomic DNA extractions of individuals A, B and C failed to control for intra-individual heterogeneity in mtDNA sequences (Table 1; Fig. 1B). Amplifications using the universal mtDNA primers on ~6000 bp template DNA obtained from long PCR resulted in mixed PCR products within individuals that varied in pairwise sequence divergence from 0 to 1.7% and in length from 931 to 937 bp. The extent of intra-individual variation among recovered clones could not be explained by *Taq* polymerase error (Table 1). Gene genealogy analysis of the 42 sequences obtained from the three different individuals revealed that some of the sequences were not monophyletic with respect to the individual genomes from which they were recovered (Fig. 1B). A BLAST nucleotide search of the highly divergent sequences shown in Fig. 1(A, B) placed them with those from other grasshoppers, indicating that they were not a result of extraneous contamination.

Examination of the variation among clones at phylogenetically conserved motifs $\geq$ 5 bp within the sequenced 12S-valine-16S region revealed that none of the most commonly recovered haplotypes within individuals following either standard or long PCR amplification contained mutations in these conserved regions (Fig. 1A, B). Conversely, the most divergent sequences that could not be reliably placed in the haplotype networks (A-4a, C-4a and A-10b in Fig. 1A, B) each contained a number of changes within the conserved motifs, including shared ancestral polymorphisms present in other taxa, but not among the other clones from *H. viridis* individuals. Similarly, a majority of the singleton haplotypes that could be reliably placed within the haplotype networks also contained unique mutations within these conserved motifs (Fig. 1). These changes in conserved regions were observed in all five of the singleton haplotypes observed following standard PCR (Fig. 1A) and in 18 of 30 singleton haplotypes observed following long PCR (Fig. 1B)

Intra-individual heterogeneity in ITS sequences

PCR amplification of the ITS1-5.8S-ITS2 region of nuclear rDNA in individuals D and E also resulted in substantial intra-individual sequence variation (Table 1; Fig. 2). Recovered clones within individuals ranged from 0 to 0.6% in pairwise sequence divergence and from 791 to 795 bp in length. The level of observed intragenomic variation could not be accounted for by *Taq* polymerase errors during PCR (Table 1). Gene genealogy analysis of the 19 sequenced clones recovered from the two individuals revealed that sequences obtained from the same individual were not monophyletic within individual genomes (Fig. 2).

Table 1. Summary of heterogeneity in 12S-valine-16S mitochondrial DNA (mtDNA) and internal transcribed spacer (ITS) nuclear ribosomal DNA sequences in individual *Hesperotettix viridis* grasshoppers.

Sequence	Individual	Template DNA	Total clones	Unique clones	Pairwise divergence (%)	Length (bp)	Changes	*P*
12S-valine-16S	A	Enriched mtDNA*	4	3	0–1.9	931–937	24	<0.0001
		mtDNA long PCR	14	12	0–1.7	931–937	43	<0.0001
	B	Whole genomic	8	4	0–0.6	934	6	0.038
		mtDNA long PCR	14	12	0–0.9	934	26	<0.0001
	C	Whole genomic	8	4	0–5.0	931–935	53	<0.0001
		mtDNA long PCR	14	10	0–1.0	933–935	19	<0.0001
ITS1-5.8S-ITS2	D	Whole genomic	10	8	0–0.6	793–794	10	<0.0001
	E	Whole genomic	9	8	0–0.5	791–795	10	<0.0001

*Tamura & Aotsuka (1988) protocol.
PCR, polymerase chain reaction.

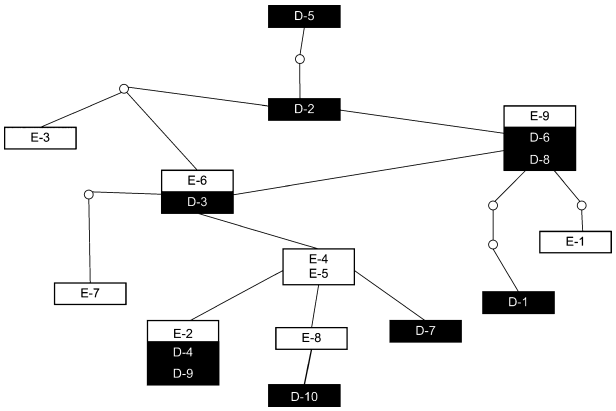

Fig. 2. Haplotype network for the ITS1-5.8S-ITS2 nuclear ribosomal DNA region for 19 sequenced clones from two individual *Hesperotettix viridis* grasshoppers (D and E). Squares represent haplotypes recovered, and circles along lineages in between squares indicate haplotypes not recovered. Each link between haplotypes indicates one mutational event, with indels, no matter what size, coded as a single event. The angle of bifurcation and the length of the link between haplotypes have no significance.

Discussion

Our results clearly demonstrate the presence in *H. viridis* grasshoppers of intra-individual variation in both 12S-valine-16S mtDNA and nuclear ITS rDNA sequences. These findings violate the core assumption underlying DNA sequence data obtained via PCR amplification for use in molecular systematics investigations (Palumbi, 1996). As described in detail elsewhere, unchecked intra-individual variation of the sort documented here can result in erroneous conclusions in phylogenetic analyses at a range of taxonomic levels (Zhang & Hewitt, 1996; Bensasson *et al*., 2001a; Álvarez & Wendel, 2003). The fact that intra-individual variation was detected in two independent DNA marker sets in the same organism only serves to strengthen the notion that the orthology of PCR-derived DNA sequence data should be thoroughly examined prior to its inclusion in molecular phylogenetic analyses. This is further reinforced by the fact that both mtDNA and nuclear rDNA ITS sequences have also been shown in separate studies to vary within individuals of a different grasshopper species, *Chorthippus parallelus* (Gomphocerinae) (Bensasson *et al*., 2000; Parkin & Butlin, 2004). These findings also constitute the first demonstration of intra-individual heterogeneity in mtDNA-like sequences in the grasshopper subfamily, Melanoplinae, and only the second report of intra-individual variation in ITS rDNA sequences in grasshoppers (Parkin & Butlin, 2004).

As a result of the difficulties reported here in establishing the orthology of 12S-valine-16S mtDNA and nuclear rDNA ITS sequences, a third independent marker set, multilocus AFLP markers, was employed to examine genetic divergence among *H. viridis* grasshoppers associated with different host plants (Sword *et al*., 2005). Using this approach, *H. viridis* was shown to exist as at least two genetically distinct host plant-associated lineages with host plant affiliation accounting for 20% of the observed genetic variation between the different lineages. In the mtDNA analyses presented here, individuals A and B were from one of the genetically distinct lineages, whereas individual C was from the other. Despite the considerable degree of genetic divergence known to exist between these individuals, the monophyly of the heterogeneous mtDNA sequences obtained from them could not always be reliably established (Fig. 1A, B).

The two individuals used in our analysis of intra-individual variation in ITS sequences (D and E) were not representatives of the different *H. viridis* lineages, so a similar contrast to that above cannot be made. Different sequences obtained from each individual were clearly not monophyletic, however, and a number of different shared polymorphisms and shared haplotypes were recovered (Fig. 2). Given that similar intra-individual variation in ITS sequences is known to occur in other grasshoppers (Parkin & Butlin, 2004), phylogenetic studies of closely related grasshopper species using these data should be carefully controlled for such variation. If intra-individual variation is present, the degree of variation among paralogous sequences can be assessed and incorporated into phylogenetic analyses (Wörheide *et al*., 2004).

With respect to mtDNA-like sequences in grasshoppers, a major source of intra-individual heterogeneity is thought to be due to the presence of numts (Bensasson *et al*., 2000). Because our attempts to isolate mtDNA were apparently unsuccessful, we are currently unable to definitively differentiate between numts and heteroplasmy as the source of the intra-individual mtDNA sequence variation observed in *H. viridis* grasshoppers. Interested readers are referred to Bensasson *et al*. (2001a) for ways in which the identity and origin of numts can be established, and their amplification potentially avoided. Examining codon position substitution bias as a means of identifying nonfunctional numt sequences (Zhang & Hewitt, 1996; Bensasson *et al*., 2000) cannot be applied to the nonprotein coding 12S-valine-16S rDNA sequences examined here. Investigating the secondary structure of transcribed RNA molecules as a means of identifying nonfunctional numts has also been suggested. However, in a critical examination of the technique, Olson & Yoder (2002) found the approach to be largely unreliable for the identification of numts.

Given the limitations in identifying putative numts in rDNA sequences, we utilized an alignment-based approach to help differentiate between potentially functional and nonfunctional mtDNA sequences by checking for polymorphism among clones in phylogentically conserved motifs. Because mtDNA sequences inserted into the nuclear genome are nonfunctional and therefore lack selective constraints, mutations accumulated in phylogentically conserved motifs may be indicative of numt status (Bensasson *et al*., 2001a,b). A majority of the most common shared haplotypes obtained from each individual following both standard and long PCR amplifications did not contain

changes in these conserved regions (Fig. 1). Although the possibility of the preferential amplification of rDNA numts cannot be ruled out on the basis of this evidence (Olson & Yoder, 2002), it remains consistent with the amplification of functional and high copy number true mtDNA sequences. On the other hand, the most divergent sequences we obtained shared a number of characteristics that suggest that they could be of nuclear origin (Fig. 1). All contained a variety of unique mutations in highly conserved motifs, indicating that they could be nonfunctional numts. Even more convincing was that these divergent sequences also retained ancestral character states found among other grasshopper species, but not in the other *H. viridis* clones. This is consistent with the notion that these sequences represent ancient nuclear translocations predating the evolution of *H. viridis*.

Interestingly, a majority of the less divergent sequences obtained also had mutations in phylogenetically conserved motifs (Fig. 1A, B). If taken as evidence that these singletons are nonfunctional numts, this finding suggests that the transfer of mtDNA to the nucleus in *H. viridis* may be an ongoing and frequent process. This possibility is further supported by the observation that the mtDNA and putative numt sequences appear to be monophyletic within individual genomes (Fig. 1A, B). However, it is important to note that the three individuals in Fig. 1(A, B) were from different allopatric populations. Thus, although mtDNA and numts would be expected to assort independently among individuals within populations, recent nuclear introgression in conjunction with restricted gene flow among populations could account for the observed pattern of intra-individual monophyly. Under this scenario, the observed monophyly within individuals of numts and mtDNA should break down with sampling of additional individuals from each population. Alternative, but seemingly less probable, explanations are either that these divergent singletons represent heteroplasmic haplotypes that are functional despite mutations in highly conserved regions, or that the *Taq* error rates experienced in this study were considerably higher than published estimates.

One of most obvious ways to avoid the potential amplification of numts is to enrich mtDNA prior to PCR amplification and sequencing (Bensasson *et al*., 2001a). Based on the assumption that numt contamination was the source of our problems in direct sequencing of 12S-valine-16S mtDNA, we tested the use of an alkaline lysis protocol for the isolation of mtDNA (Tamura & Aotsuka, 1988) on individual A in order to enrich mtDNA and potentially control for the presence of numts prior to standard PCR. This commonly used procedure is a variation of a protocol used to isolate plasmid DNA and has been successfully employed to enrich mtDNA in organisms known to harbour numts (Williams & Knowlton, 2001). Despite this, one of the four sequenced clones from individual A was highly divergent from the other A clones, exhibiting up to 1.9% sequence divergence and containing four changes in highly conserved motifs (Table 1; Fig. 1A). The finding of a sequence probably of nuclear origin in one of four clones indicates that our attempt to enrich mtDNA was largely unsuccessful. Given that the Tamura & Aotsuka (1988) protocol has been used successfully for similar purposes in other taxa (Williams & Knowlton, 2001), experimental error appears to be the most probable explanation for our results in this case.

We also employed a nested long PCR approach in an attempt to control for possible numt contamination, as suggested by Roehrdanz & Degrugillier (1998) and utilized by Bensasson *et al*. (2000) [note: Bensasson *et al*. (2000) used enriched mtDNA as the initial long PCR template]. We first amplified an ~6000 bp portion of the grasshopper mitochondrial genome. This PCR product was then used as template DNA in a subsequent PCR amplifying the smaller 12S-valine-16S mtDNA target region. Even the application of this technique failed to yield homogeneous PCR products and substantial variation was still observed within individuals, including many haplotypes that contained unique mutations in highly conserved regions (Table 1; Fig. 1B). Assuming that these variant clones are nonfunctional numts as opposed to functional heteroplasmic haplotypes, these results suggest that a considerable number of > 6000 bp numts are present in the nuclear genome of *H. viridis* grasshoppers. This may seem unlikely, but large nuclear insertions of multiple copy number are known to occur in other organisms, with one of the most notable examples being in cats, where a 7.9 kb numt is present in the nuclear genome as 38–78 tandemly repeated copies (Lopez *et al*., 1994). In addition, a numt in humans is known to consist of 88% of the entire 16.5 kb mitochondrial genome (Tourmen *et al*., 2002).

Our results elucidate a major potential pitfall in collecting DNA sequence data via PCR for use in molecular phylogenetic analyses. Extreme care should be taken to ensure that the presence of intra-individual variation is assessed and that the orthology of compared sequences across taxa can be assured. Our findings are particularly relevant to organisms, such as grasshoppers, in which intragenomic variation is known to be an issue (Zhang & Hewitt, 1996; Bensasson *et al*., 2000; Parkin & Butlin, 2004). Acridid grasshoppers have been used in many mtDNA-based phylogenetic analyses (Chapco *et al*., 1997, 1999; Flook & Rowell, 1997a, b; Flook *et al*., 2000; Knowles & Otte, 2000; Dopman *et al*., 2002; Litzenberger & Chapco, 2003), but only recently have grasshopper molecular systematics investigations begun to incorporate explicit controls for the presence of intra-individual sequence heterogeneity such as the clone and sequence approach employed here (Lovejoy *et al*., 2006). Importantly, neither the presence nor the full extent of variation among different sequences within an individual is necessarily evident from the examination of chromatograms obtained when sequencing directly from PCR products. In our study, the specific polymorphisms found among different clones from the same individual were typically not evident and did not directly correspond to difficult to resolve regions of sequence chromatograms generated when sequencing directly from the same PCR product. This was presumably due in part to length heterogeneity among the

different sequences, as well as differences in their relative frequency following PCR amplification.

The findings presented here of yet another example of intra-individual DNA sequence variation in grasshoppers also highlights the need for caution in the use of DNA barcoding for species taxonomy (Hebert *et al*., 2003). The inadvertent analysis of paralogous DNA sequences is known to be a weakness of DNA barcoding (Hebert *et al*., 2003; Tautz *et al*., 2003) and can limit the utility of such an approach for use in taxonomy and species identification (Moritz & Cicero, 2004; Thalmann *et al*., 2004; Pons & Volger, 2005).

Acknowledgements

Douda Bensasson kindly provided discussion and advice early in this project regarding the detection and avoidance of numts. Roger Butlin, Jenny Apple and an anonymous reviewer provided helpful comments on the manuscript. Thanks to Robert Lartey for his expertise and guidance in the laboratory. Benjamin Duval assisted with the genomic DNA extractions. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

References

Álvarez, I. & Wendel, J.F. (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, **29**, 417–434.

Avise, J.C. (2004) *Molecular Markers, Natural History and Evolution*. Sinauer Associates, Sunderland, Massachusetts.

Ballard, J.W.O. & Whitlock, M.C. (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Bensasson, D., Petrov, D.A., Zhang, D.X., Hartl, D.L. & Hewitt, G.M. (2001b) Genomic gigantism: DNA loss is slow in mountain grasshoppers. *Molecular Biology and Evolution*, **18**, 246–253.

Bensasson, D., Zhang, D.-X., Hartl, D. & Hewitt, G.M. (2001a) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution*, **16**, 314–321.

Bensasson, D., Zhang, D.X. & Hewitt, G.M. (2000) Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Molecular Biology and Evolution*, **17**, 406–415.

Chapco, W., Kuperus, W.R. & Litzenberger, G. (1999) Molecular phylogeny of melanopline grasshoppers (Orthoptera: Acrididae): the genus *Melanoplus*. *Annals of the Entomological Society of America*, **92**, 617–623.

Chapco, W., Martel, R.K.B. & Kuperus, W.R. (1997) Molecular phylogeny of North American band-winged grasshoppers (Orthoptera: Acrididae). *Annals of the Entomological Society of America*, **90**, 555–562.

Clement, M., Posada, D. & Crandall, K. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Dopman, E.B., Sword, G.A. & Hillis, D.M. (2002) The importance of the ontogenetic niche in resource-associated divergence: evidence from a generalist grasshopper. *Evolution*, **56**, 731–740.

Dowling, T.E., Moritz, C., Palmer, J.D. & Riesberg, L.H. (1996) Nucleic acids III: analysis of fragments and restriction sites. *Molecular Systematics* (ed. by D.M. Hillis, *et al*.), pp. 249–320. Sinauer Associates, Sunderland, Massachusetts.

Flook, P.K., Klee, S. & Rowell, C.H.F. (2000) Molecular phylogenetic analysis of the Pneumoroidea (Orthoptera, Caelifera): molecular data resolve morphological character conflicts in the basal Acridomorpha. *Molecular Phylogenetics and Evolution*, **15**, 345–354.

Flook, P.K. & Rowell, C.H.F. (1997a) The phylogeny of the Caelifera (Insecta, Orthoptera) as deduced from mtrRNA gene sequences. *Molecular Phylogenetics and Evolution*, **8**, 89–103.

Flook, P.K. & Rowell, C.H.F. (1997b) The effectiveness of mitochondrial rRNA gene sequences for the reconstruction of the phylogeny of an insect order (Orthoptera). *Molecular Phylogenetics and Evolution*, **8**, 177–192.

Flook, P.K., Rowell, C.H.F. & Gellissen, G. (1995) The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. *Journal of Molecular Evolution*, **41**, 928–941.

Funk, D.J. & Omland, K.E. (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 397–423.

Hebert, P.D.N., Cywinska, A., Ball, S. & deWaard, J. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society Biology Sciences Series B*, **270**, 313–321.

Hillis, D.M. & Dixon, M.T. (1991) Ribosomal DNA – molecular evolution and phylogenetic inference. *Quarterly Review of Biology*, **66**, 410–453.

Hillis, D.M., Moritz, C. & Mable, B.K. (1996) *Molecular Systematics*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Knowles, L.L. & Otte, D. (2000) Phylogenetic analysis of montane grasshoppers from western North America (genus *Melanoplus*, Acrididae: Melanoplinae). *Annals of the Entomological Society of America*, **93**, 421–431.

Kobayashi, N., Tamura, K. & Aotsuka, T. (1999) PCR error and molecular population genetics. *Biochemical Genetics*, **37**, 317–321.

Leo, N.P. & Barker, S.C. (2002) Intragenomic variation in ITS2 rDNA in the louse of humans, *Pediculus humanus*: ITS2 is not a suitable marker for population studies in this species. *Insect Molecular Biology*, **11**, 651–657.

Liao, D. (1999) Concerted evolution: molecular mechanism and biological implications. *American Journal of Human Genetics*, **64**, 24–30.

Litzenberger, G. & Chapco, W. (2003) The North American Melanoplinae (Orthoptera: Acrididae): a molecular phylogenetic study of their origins and taxonomic relationships. *Annals of the Entomological Society of America*, **96**, 491–497.

Lopez, J.V., Yuhki, N., Masuda, R., Modi, W. & O'Brien, S.J.O. (1994) *Numt*, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution*, **39**, 174–190.

Lovejoy, N.R., Mullen, S.B., Sword, G.A., Chapman, R.F. & Harrison, R. (2006) Ancient trans-Atlantic flight explains locust biogeography: molecular phylogenetics of *Schistocerca*. *Proceedings of the Royal Society of London Series B*, **273**, 767–774.

Loxdale, H.D. & Lushai, G. (1998) Molecular markers in entomology. *Bulletin of Entomological Research*, **88**, 577–600.

Moritz, C. & Cicero, C. (2004) DNA barcoding: promise and pitfalls. *Public Library of Science Biology*, **2**, 1529–1531.

Olson, L.E. & Yoder, A.D. (2002) Using secondary structure to identify ribosomal numts: cautionary examples from the human genome. *Molecular Biology and Evolution*, **19**, 93–100.

Palumbi, S.R. (1996) Nucleic acids II: the polymerase chain reaction. *Molecular Systematics* (ed. by D.M. Hillis, *et al.*), pp. 205–248. Sinauer Associates, Sunderland, Massachusetts.

Parkin, E.J. & Butlin, R.K. (2004) Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intrachromosomal gene conversion. *Molecular Biology and Evolution*, **21**, 1595–1601.

Pons, J. & Volger, A.P. (2005) Complex pattern of coalescence and fast evolution of a mitochondrial rRNA pseudogene in a recent radiation of tiger beetles. *Molecular Biology and Evolution*, **22**, 991–1000.

Richly, E. & Leister, D. (2004) NUMTs in sequenced eukaryotic genomes. *Molecular Biology and Evolution*, **21**, 1081–1084.

Roehrdanz, R.L. & Degrugillier, M.E. (1998) Long sections of mitochondrial DNA amplified from fourteen orders of insects using conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **91**, 771–778.

Rokas, A., Ladoukakis, E. & Zouros, E. (2003) Animal mitochondrial DNA recombination revisited. *Trends in Ecology and Evolution*, **18**, 411–417.

Sanderson, M.J. & Shaffer, H.B. (2002) Troubleshooting molecular phylogenetic analyses. *Annual Review of Ecology and Systematics*, **33**, 49–72.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution, weighting and the phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Sokal, R.R. & Rohlf, F.J. (1995) *Biometry*, 3rd edn. W.H. Freeman, New York.

Sorenson, M.D. & Quinn, T.W. (1998) Numts: a challenge for avian systematics and population biology. *Auk*, **115**, 214–221.

Swofford, D.L. (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, Massachusetts.

Sword, G.A., Joern, A. & Senior, L.B. (2005) Host plant-associated genetic differentiation in the snakeweed grasshopper, *Hesperotettix viridis* (Orthoptera: Acrididae). *Molecular Ecology*, **14**, 2197–2205.

Tamura, K. & Aotsuka, T. (1988) Rapid isolation method of animal mitochondrial-DNA by the alkaline lysis procedure. *Biochemical Genetics*, **26**, 815–819.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. & Vogler, A.P. (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution*, **18**, 70–74.

Templeton, A.R., Crandall, K.A. & Sing, C.F. (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

Thalmann, O., Hebler, J., Poinar, H.N., Pääbo, S. & Vigilant, L. (2004) Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. *Molecular Ecology*, **13**, 321–335.

Tourmen, Y., Baris, O., Dessen, P., Jacques, C., Malthiery, Y. & Reynier, P. (2002) Structure and chromosomal distribution of human mitochondrial pseudogenes. *Genomics*, **80**, 71–77.

Vogler, A.P. & Desalle, R. (1994) Evolution and phylogenetic information content of the ITS-1 region in the tiger beetle *Cicindela dorsalis*. *Molecular Biology and Evolution*, **11**, 393–405.

Weekers, P.H.H., De Jonckheere, J.F. & Dumont, H.J. (2001) Phylogenetic relationships inferred from ribosomal ITS sequences and biogeographic patterns in representatives of the genus *Calopteryx* (Insecta: Odonata) of the west Mediterranean and adjacent West European zone. *Molecular Phylogenetics and Evolution*, **20**, 89–99.

Williams, S.T. & Knowlton, N. (2001) Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus *Alpheus*. *Molecular Biology and Evolution*, **18**, 1484–1493.

Wörheide, G., Nichols, S.A. & Goldberg, J. (2004) Intragenomic variation of the rDNA internal transcribed spacers in sponges (phylum Porifera): implications for phylogenetic studies. *Molecular Phylogenetics and Evolution*, **33**, 816–830.

Zhang, D.-X. & Hewitt, G.M. (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution*, **11**, 247–251.

Accepted 18 January 2007
First published online 14 May 2007

